

CHAPTER 8: SCALE SCORES & PROFICIENCY LEVELS

RATIONALE

To ensure that student proficiency results for each grade are reported on a common standard score scale, EED provides a unique scale score conversion system for each SBA assessment. In this system, raw scores are converted to a logistic metric. Logit measures are then transformed into scale scores. Scale scores are intended to make scores more meaningful by defining a scale of measurement that is not tied to a particular test form. The scale scores across all subjects and grades are identical, with a maximum of 600, a minimum of 100, and the basic/proficient cut score set to 300.

DESCRIPTION OF SCORES

Raw Score

The basic summary statistic on all SBA assessments is the raw score. A raw score is reported for each examinee in mathematics, reading, and writing. The raw score is the number of multiple-choice items answered correctly plus the number of points earned on constructed-response items. By itself, the raw score has limited utility; it can only be interpreted in reference to the total number of items on a subject-area assessment. That is, receiving a raw score of 30 points out of a total possible of 40 has no meaning without some knowledge of the difficulty of the test.

Scale Score

To be fair to students and to permit valid comparison of test scores across administrations, the skills represented by each score point must remain consistent from year to year. Scale scores are statistical conversions of raw scores that adjust for differences in item and form difficulties and permit valid comparisons across test administration. Scale scores fall into a category referred to as interval-level scales. Scale scores, in contrast to raw scores, allow for the interpretation of the scores to be the same at any point along the scale continuum. For example, the interpretation of a ten-point difference between scores of 260 and 270 is the same as that for scores of 390 and 400. With scale scores, schools can compare the demonstrated knowledge and ability of groups of students across years. Comparing scale scores on the assessments can also help schools determine the impact of instruction and curriculum. Note that the scale score range for the SBA is from a minimum of 100 to a maximum of 600.

SCALING AND PROFICIENCY LEVELS

The SBA uses a unique scale score conversion system for each assessment. In this system, raw scores are first converted to a logistic metric. Thereafter, logit measures are transformed into scale scores, the final reporting metric. To assist in making meaningful interpretations, the Proficient cutpoint is fixed at a scale score of 300 for all grades and subjects and becomes the center of origin of each scale. To determine the location of the scale score, a multiplier is applied to the logit that represents the standard deviation of the student scores. The formulas used for each grade and subject are provided in Tables 8-4 to 8-6 of the 2005 *Standards Validation Report*.

Standards Validation

The composition of this panel was twenty-four teachers and non-teacher educators for each of the three subjects (reading, math, and writing for a total of 72. Within each subject, all panelists participated in judgments for grade 6. Thereafter, two separate groups were formed. One group conducted grades 5, 4, and 3 and another conducted grades 7, 8, and 9.

A national technical advisory committee (TAC), convened in fall 2004 and again in spring 2005, recommended the use of the existing grade 6 cutpoints (by subject) at the starting point for all grades. The rationale for this recommendation was based on a combination of the trends in the data, the history of the program, preparing for AYP, and the interaction between the results of the benchmark tests in grades 3, 6, and 8 and the norm-referenced tests given in the gap grades.

The starting points were established by setting the percent of students that would be placed into each of the four proficiency levels to be equal for all grades. During the course of the meeting, participants were allowed to make recommendations that varied from these initial guidelines, as long as they had substantive reasons for doing so. The participants were instructed to either validate or suggest new cutpoints. In effect, they went into the process with the goal of articulating the cutpoints across grades. In this sense, their charge was to establish cut scores across grades that would produce comparable levels of academic achievement.

As a final step, the panelists' recommendations from each subject were adjusted via a postsMOOTHING process. This process used a non-parametric tri-weight kernel smoother with polynomial degree two¹. The goal was to achieve a smoother progression across the grades with appropriate consideration of weight given to the panelists' recommendations. To achieve this goal, the smoothing function applied to the data was selected in consultation with the TAC and department staff.

Comparability of Scale Scores Across Grades

Through the process described in the previous section, the standards for Proficient were established to have consistent interpretation from grade to grade. The logit measures that defined the cut score for each subject and grade (obtained from the standard setting process described above) was then defined to be a scale score of 300. Thus, a student who receives a scale score of 300 at each grade is making progress from grade to grade that exactly equals the difference in the standards for Proficient across those two grades.

Further, the relationship between the logit measure and the scale scores was established so that the standard deviation of scale scores would be 75 on average across all the grades. Because the standard deviation of the logit measure varied slightly from grade to grade, the standard deviation of student scale scores is slightly higher than 75 at some grades and slightly less than that amount at others, but for all grades and subject, the standard deviation of the scale scores is approximately 75 (they range from a low of 73.6 for grade 8 writing to a high of 75.3 for grade 5 reading).

As a result, the interpretation of scale scores is the same for all grades and subjects in the following context: a scale score of 225, for example, means that the student scored approximately one standard deviation below the standard for Proficient. If that same student had a scale score of 250 in that subject at the next grade (meaning the student now is approximately just .67 standard deviations below the standard for Proficient), the student is now closer to the standard of Proficient at this grade than he/she was the year previously to the standard for Proficient at the lower grade. Restated, a higher scale score at one grade than another means that the student is achieving better relative to the standard for Proficient at the higher grade.

GRADE 10 STANDARDS SETTING

Traditionally, standard setting methods have fallen into two camps: test-centered methods and examinee-centered methods (Jaeger, 1989). With test-centered methods, the standard-setting judgment is made primarily by referencing the test itself, most often based upon an inspection of the

¹ SPSS Inc. 1999. SYSTAT 9: Statistics II. Chicago: SPSS Inc.

actual test items. Examinee-centered methods call for judgments to be made about the performance of examinees, such as the performance of established mastery and non-mastery groups. As Kane (1995) points out, “all standard setting is based on judgments.” Performance standard setting uses a process from which the best judgments are obtained from the people in the best position to make those judgments. Typically, these are content experts, people familiar with the skills and knowledge to be learned.

Standard Setting Panels—Alaska Teachers and Stakeholders

Representatives of three broad groups typically comprise the standard setting panel. These groups include teachers, non-teacher educators, and non-educators. Non-educators are drawn from civic and business leaders. The non-teacher educators could be curriculum directors or administrators. Its selection process also considers size and location of districts and schools, socioeconomic conditions, and other demographics.

While there is no simple answer to the question of how many panelists are required to establish reliable standards, the larger the number involved the more confidence there can be that the result will generalize to another selection of panelists. The current plan sought 15 panelists for each content area or a total of 45. Separate panels were used for mathematics, reading, and writing.

Modified Bookmark Procedures

Because there was a need to maintain consistency with grades 3–9, the same Modified Bookmark Procedure (Lewis, Mitzel, Green, & Patz, 1999) was utilized to set the cut scores. A schedule for the May 2006 Standard Setting is provided in Appendix 21.

Grade 10 Recommended Cut Scores

Appendix 22 shows the grade 10 round by round summaries for each content area. Once the panelists’ recommendation was calculated, the impact data was appended to the grades 3–9 impacts within each subject area.

TRANSFORMATIONS

The student ability measures were transformed mathematically to a more convenient metric. To maintain consistency from administration to administration, the minimum scale scores necessary for each proficiency level are provided in Tables 8–1 through 8–3. Tables 8–4 through 8–6 provide the equations used for each transformation. These equations were applied to the overall test as well as to each reporting subscale.

Table 8–1. Minimum Mathematics Scale Scores for Each Proficiency Level

Mathematics

Grade	Raw Score Cut Point			Below Proficient		Proficient		Advanced	
	Below Proficient	Proficient	Advanced	SS Cut	SSSE	SS Cut	SSSE	SS Cut	SSSE
3	25	33	51	263	17	300	17	390	20
4	27	35	51	260	18	300	18	383	22
5	25	35	50	252	17	300	17	373	20
6	25	34	49	258	17	300	17	376	19
7	22	33	49	248	18	300	17	383	20
8	25	34	50	258	18	300	17	379	20
9	26	36	50	258	17	300	17	370	20
10	16	22	33	252	22	300	21	392	26

Table 8–2. Minimum Reading Scale Scores for Each Proficiency Level**Reading**

Grade	Raw Score Cut Point			Below Proficient		Proficient		Advanced	
	Below Proficient	Proficient	Advanced	SS Cut	SSSE	SS Cut	SSSE	SS Cut	SSSE
3	18	26	46	261	17	300	16	392	19
4	19	27	48	260	19	300	18	415	22
5	16	25	48	251	19	300	17	418	21
6	17	29	47	234	19	300	17	394	21
7	18	27	46	246	20	300	19	406	22
8	17	27	46	243	20	300	18	402	21
9	15	28	45	229	20	300	18	382	20
10	28	45	62	222	18	300	18	400	25

Table 8–3. Minimum Writing Scale Scores for Each Proficiency Level**Writing**

Grade	Raw Score Cut Point			Below Proficient		Proficient		Advanced	
	Below Proficient	Proficient	Advanced	SS Cut	SSSE	SS Cut	SSSE	SS Cut	SSSE
3	12	27	48	218	20	300	17	402	21
4	13	27	46	204	25	300	21	420	24
5	12	31	49	187	23	300	19	406	25
6	17	32	47	215	22	300	20	396	25
7	20	32	51	234	19	300	18	423	26
8	19	31	53	232	20	300	19	460	30
9	21	33	55	238	20	300	19	470	35
10	21	31	46	233	20	300	23	485	42

Table 8–4. Equations Used for Each Transformation in Mathematics**Mathematics**

Grade	Conversion Equation	Logit Cuts		
		BP	P	A
3	Scale Score = (61.9835 x Logit) + 291.0799	-0.4508	0.1358	1.6023
4	Scale Score = (66.9643 x Logit) + 280.6843	-0.3091	0.2810	1.5209
5	Scale Score = (61.9835 x Logit) + 283.6666	-0.5175	0.2554	1.4354
6	Scale Score = (63.0252 x Logit) + 282.7344	-0.3923	0.2660	1.4826
7	Scale Score = (64.1026 x Logit) + 288.3632	-0.6263	0.1737	1.4709
8	Scale Score = (63.5593 x Logit) + 277.1580	-0.3074	0.3515	1.6095
9	Scale Score = (63.0252 x Logit) + 275.7762	-0.2816	0.3764	1.5009
10	Scale Score = (58.1395 x Logit) + 274.5000	-0.3861	0.4300	2.0123